DOCUMENT RESUME

ED 285 737 SE 048 325

AUTHOR Moore, Joyce L.
TITLE Back-of-the-Envelope Problems. Final Report.
INSTITUTION California Univ., Berkeley.
SPONS AGENCY Office of Naval Research, Washington, D.C. Personnel and Training Branch.
REPORT NO GK-3
PUB DATE Jul 87
CONTRACT N00014-85-K-0095
NOTE 53p.; For related documents, see SE 048 323-327.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Structures; Engineering; *Estimation (Mathematics); *Inferences; Learning Strategies; Mathematical Applications; *Problem Solving; Science Education; *Sciences
IDENTIFIERS *Experts

ABSTRACT

Back-of-the-envelope problems call for approximate calculations of quantities that can be related to information in a person's knowledge but are not solved precisely. These problems provide an opportunity for the study of processes and the role of general knowledge in ill-defined problem solving. Subjects with advanced and intermediate knowledge in different domains provided protocols solving back-of-the-envelope problems within and outside their fields of special knowledge. Their protocols were interpreted using a set of distinctions provided by Schoenfeld (1985). A model that simulates important aspects of the observed performance was developed by extending FERMI, a model of general problem-solving methods by Larkin, Reif, Carbonell, and Gugliatta (1985). An important feature of expert performance was the ability to use domain-specific knowledge in the service of general problem-solving methods. General methods of problem solving were similar across subjects who differed in knowledge. Expert knowledge probably provided guidance in the choice of general methods, as well as relevant specific knowledge for solutions. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

SE

July 1987

Report No. GK-3

ED285737

Back-of-the-Envelope Problems

Joyce L. Moore

University of California, Berkeley

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

This research was funded by the Office of Naval Research with Contract N00014-85-K-0095, Project NR 667-544. Approved for public release; distribution unlimited.

SE 048 335

BEST COPY AVAILABLE

Back-of-the-Envelope Problems

Joyce L. Moore

University of California, Berkeley

ABSTRACT

Back-of-the-envelope problems call for approximate calculations of quantities that can be related to information in a person's knowledge but are not solved precisely. They therefore provide an opportunity for study of processes and role of general knowledge in ill-defined problem solving. Subjects with advanced and intermediate knowledge in different domains provided protocols solving back-of-the-envelope problems within and outside their fields of special knowledge. Their protocols were interpreted using a set of distinctions provided by Schoenfeld (1985). A model that simulates important aspects of the observed performance was developed by extending FERMI, a model of general problem-solving methods by Larkin, Reif, Carbonell, and Gugliatta (1985). An important feature of expert performance was the ability to use domain-specific knowledge in the service of general problem-solving methods. General methods of problem solving were similar across subjects who differed in knowledge. At the same time, expert knowledge probably provided guidance in the choice of general methods, as well as relevant specific knowledge for solutions.

This research was supported by the Office of Naval Research, Contract N00014-85-K-0095, Project NR 667-544.

Back-of-the-Envelope Problems

Scientists can occasionally be found scribbling calculations for 'back–of–the–envelope' problems (BEP). These are rough computations often performed on envelopes or other scraps of paper. Usually these problems, sometimes called 'order–of-magnitude' problems, involve a series of estimations. Several well-known examples, attributed to Enrico Fermi, are "How many piano tuners are there in New York City?" and "How much does a watch gain or lose when carried up a mountain?" A more practical type of problem, though in a similar vein, is encountered by an engineer who computes a rough estimate to determine the feasibility of a proposed design.

The arithmetic involved in back-of-the-envelope calculations is usually quite simple. The difficult part seems to be retrieving the necessary facts from memory or estimating them in some reasonable way. It is also necessary to know where to make rough estimates and where more accurate data is needed in order for the calculation to be useful. This information should also feed into a hypothesis of the probable magnitude of error in the final answer.

Morrison (1963) feels that these questions draw upon a deep understanding of the world, everyday experience, and the ability to make rough approximations, inspired guesses, and statistical estimates from very little data. The skill derived from answering this type of question is proffered as good apprenticeship to research. Morrison suggests that back-of-the-envelope problems cultivate an ability which is as valuable as the more formal sort gained from standard classroom instruction. In addition, he feels that back-of-the-envelope problems of varying difficulty can be used at many levels of education.

Back-of-the-envelope problems seem to be part of the culture of several disciplines and are intuitively felt to provide a valuable skill. This belief, however, has remained intuitive. The

reasons why back-of-the-envelope problems foster competent problem solving, or if indeed they do, has remained unexplored. As a first step in the investigation of this issue, the structure of these problems and the processes used to solve them will be analyzed. A model of expert solution processes on these problems will then be proposed. This model will be used to support inferences about the organization of necessary expert knowledge structures. This should eventually allow us to explore and predict the effect of different knowledge organizations on problem solving performance. In the first section of this paper, the use of back-of-the-envelope problems by practicing scientists and in the instruction of students will be discussed. Second, the structure of the problems themselves will be examined and compared to the typical problems used to study expertise in several domains. Third, protocols of several experts and intermediates from different domains solving back–of–the–envelope problems will be discussed. In the final section, the model of expert solution processes will be presented.

Uses of BEP

In some domains, physics and engineering in particular, this type of quick calculation has been recognized as a critical part of the trade. Engineers will often perform rough feasibility estimates before investing a large amount of time in a project or design. Indeed, this technique is often taught as part of the standard engineering curricula (Bentley, 1984). The field of physics also recognizes the value of this type of calculation for both experienced scientists and students. The *American Journal of Physics* ran a department called "Back-of-the-Envelope" from 1983 to 1984. Each month, three questions were posed, with the answers supplied the following month. An example from the column is "How big an asteroid could you escape from by jumping?" (Purcell, 1983). The editor of the column, Edward Purcell, had often used this type of problem as an introduction to a graduate seminar in physics (personal communication, 1981). Students were expected to be able to solve the problems using only a one-page "Round Number Handbook of Physics," a list of quantities such as constants and masses, and the student's own general knowledge.

Not only is the ability to solve back-of-the-envelope-problems considered an integral aspect of expert behavior, these problems are sometimes used as a technique for predicting success. Order-of-magnitude problems have been used on tests to determine eligibility for physics programs at the high school level (A. diSessa, personal communication, 1986). For example, "How far can a goose fly?" and "How long a line can you write with a ballpoint pen?" were two of the questions on a take–home entry examination. Several reasons were given for including this type of question. First, it was a way of introducing the students to the fact that they possess a common sense knowledge which can be combined in unusual ways to answer questions which initially sound intractable. Second, it presented a method of reasoning about the world which, although quantitatively imprecise, is often sufficient given the particular question asked. Third, the questions seemed to identify students who were highly motivated to learn and understand material. On the ballpoint pen question better answers involved measuring, recalling the last time a pen lasted, how much one wrote with it, etc., while grade-oriented students produced an "academic" answer such as "You can't draw a 'perfect' line with a ballpoint pen." As with Morrison, it was felt that the answers to these questions indicated those students who would perform well in a research setting.

Recognition of the usefulness, and often necessity, of this type of calculation has more recently spread to the area of computer science. *Communications of the ACM* has recently run several columns on computer science back-of-the-envelope problems (Bentley, 1984, 1986). One problem posed in this area was "Suppose the world slowed down by a factor of a million. How long does it take for your computer to execute an instruction? Your disk to rotate once? Your disk arm to seek across the disk? You to type your name?" Bentley suggests that a few envelopes worth of arithmetic early in the life of a software project may help a system designer make rational choices and avoid a project doomed to failure.

In summary, engineers recognize the need for back-of-the-envelope calculations and often include them as part of the curriculum for training students. Physicists also recognize the usefulness of these computations and, on an individual basis, will sometimes include them in a classroom setting. The field of computer science has recently come to realize the value of back–of–the–envelope problems, though there is no evidence that they have been incorporated into the classroom.

The ability to perform back-of-the-envelope calculations competently is thus recognized in several disciplines as being an indicator, cultivator, and predictor of expertise. This recognition is largely the result of the pragmatics and demands of these fields. Engineering do not want to design structures which will require twice the money allotted. Similarly, computer scientists do not want to propose systems that would require 120 seconds in each minute. With a few exceptions, these problems seem to be used mainly in the practice of a discipline, rather than in an educational setting. Is there a basis for the use of these problems as an educational tool? The next section will examine the structure of back–of–the–envelope problems and compare them to the traditional problems of other disciplines. This may help identify those aspects of the problems actually tap expertise.

The Structure of BEP

Back-of-the-envelope problems fall into the category of ill-structured problems, though subproblems in the solution can be well-structured. Well-structured problems are those in which the initial situation, the goal state, and the operators for transforming the current state, are clearly delineated and well-defined (Greeno & Simon, in press). Traditional physics and geometry problems fall into this category and are relatively well studied (e.g., Chi, Feltovich, & Glaser, 1981; Greeno, 1978; Larkin, McDermott, Simon, & Simon, 1980). In a geometry proof, for example, the initial state consists of the given part of the proof and the desired goal state is what one must prove. The operators are the corollaries, definitions, theorems, etc. that one

uses as reasons for a step in the proof. the operators move the geometry proof from the initial state to the goal state.

One characteristic of well-structured problems is that there tends to be large classes of the problems which have agreed upon solutions. A substantial part of the solution process for an experienced problem solver therefore lies in identifying the type of the problem. As shown by Larkin et al. (1980), the physics expert spends much time creating a representation of the problem, i.e., determining the class of which the specific problem is a member. The steps to the solution are then almost trivial because the solution itself is clearly defined. In other words, after the creation of a representation the solution steps are somewhat automatic in that they are not explicitly or individually evaluated. Another factor leading to consensus among experts on physics problems is that the constraints of the problems studied have generally been well-defined and hence tend to produce agreement (Voss, Greene, Post, & Penner, 1983).

Poorly-structured problems have been less extensively researched (Reitman, 1965; Simon, 1973). A type of ill-structured problem occurs when the goals of the problem are undetermined. In well-structured problems, the goals are specific objects, such as a geometry proof statement, whereas in ill-structured problems, the undetermined goals allow for alternative solution paths (Greeno & Simon, in press). For example, writing an essay or painting a picture are both ill-structured problems. Another type of ill-structured problem occurs when the solution requires knowledge from several different sources. This necessitates the coordination of work in several disparate problem spaces (Simon, 1973). A form of this type of problem occurs in geometry problems that require the construction of auxiliary lines. Here the problem space that is given must be augmented with an operator for the construction of an auxiliary line in order for the problem to be solved.

An interesting study of ill-structured problems involve the social sciences, an area in which problems are generally ill-defined (Voss et al., 1983). An example of the problems they examined was: "Assume you are the Head of the Soviet Ministry of Agriculture, and assume crop productivity has been low over the past several years. You now have the responsibility to increase crop production. How would you go about doing this?" They noted that, unlike physics, few problems of this nature have agreed upon solutions. Because it is difficult to determine whether or not a given answer is a viable solution, or to implement an answer, social science answers are judged on the merits of supporting argument. This suggests that it is the underdetermination of social science goals that makes this argument evaluation necessary.

Another aspect of these problems which contributes to widely varying answers is that the constraints are multiple, and any given expert typically cannot or does not consider all of the possible constraints. It is also necessary for the social science expert to use real-world knowledge to determine how particular constraints operate (Voss, Tyler, & Yengo, in press). This real-world knowledge and the use of it may vary among experts creating diverse solutions.

Most back-of-the-envelope problems belong in the class of ill-structured problems. There is no agreed upon solution to many of the problems, though the form of the answer is defined. For instance, in solving the problem "How many leaves fall in North America every autumn?" one knows that the answer must take the form of a number, which represents a quantity of leaves. However, there is no 'correct' answer to this problem. Furthermore, there is no preferred solution path among solvers as many equally viable methods could be used. For example, the number of leaves per tree and the number of trees in North America could be calculated; however, the volume of leaves or the area of leaves could also provide the basis for an answer.

Because it is often impossible to judge the veridicality of a given answer, such as in the leaves problem, problem solvers tend to evaluate the process by which they arrived at their solutions. For example, if a number has been estimated the solver evaluates the procedure used to generate the answer. They attempt to justify their answer to each component or subgoal in the problem, doing this for each segment of the problem. Presumably, this provides an overall judgement of the acceptableness of the final answer, since the solver cannot check whether their answer is correct. This is similar to the process Voss et al. observed in social science experts during the generation of their answers. As each leg of the solution was outlined the expert presented a supporting argument, and then critiqued the argument. It appears that this evaluation of the solution arises from the lack of an agreed upon answer or method for testing the validity of an answer.

Since it is the nature of back-of-the-envelope problems to have no truly 'right' answer, or even a preferred solution path, this opens up many creative avenues to a solution. If the problem solver needs a necessary piece of information which s/he does not possess they are free to bypass this obstacle by estimating the necessary quantity. If it is a procedure which has been forgotten, or never learned, the problem solver can often arrive at a reasonable solution by another solution path entirely, such as a series of estimations. Often, this option is not available with typical classroom problems. When problems are asked in a classroom there is usually a single solution procedure and answer desired. Even when there are several different solution paths available they must all converge on the same answer. The open-endedness of the back of an envelope provides an escape from the strict structure of most problem sets and may encourage original, creative problem solving on the part of the students.

Despite many ill-structured aspects of back-of-the-envelope problems, well–structured domain specific methods are often necessary or useful for their solution. In those problems involving physics, such as the asteroid problem mentioned above, knowledge and formulas

about concepts like force and mass can be applied in much the same manner as in a classic textbook physics problem. However, these methods will usually supply only subparts to the overall question as they are embedded in a larger ill–structured context. It is unlikely that back-of-the-envelope problems can be classified and solved by using only well-structured methods, as with many textbook physics problems.

Perhaps the most important characteristic of back-of-the-envelope problems is their integration of both domain specific and general knowledge. This may be why Enrico Fermi found such delight in devising and answering this type of question. This may also explain why these problems have long been intuitively believed to tap more than just rote classroom learning. Even when textbook material has been well learned, back-of-the-envelope problems usually require going a step beyond this type of knowledge by applying some common sense knowledge in an unusual way. This use of the domain–specific and general knowledge, and consequently domain–specific and general knowledge, and consequently domain-specific and general methods, requires work in several different problem spaces. The coordination of knowledge from separate sources may add a layer of difficulty to the problems. It does provide another ill–structured aspect of this type of problem.

In summary, there are several factors which combine to make back-of-the envelope problems challenging. First, they require the retrieval and organization of several different types of knowledge; domain specific and general. Second, the problems combine both the creativity invited by ill-structured problems with the analytic skills necessary to solve well-structured subproblems. Third, the lack of an agreed upon answer often invokes a process of evaluation of one's problem solving processes in order to judge the acceptableness of a solution. Back-of-the-envelope problems may provide a valuable instructional tool for these reasons.

BEP Protocols

In this section, protocols of several experts and intermediates solving back-of-the-envelope problems will be discussed. This provides examples of both the structured and ill-structured aspects of these problems. In addition, there will be evidence of the use of domain specific and general problem solving methods.

A pilot study was conducted in which two factors affecting the solution of back-of-the-envelope problems were studied. The level of expertise of the problem solvers was varied to observe any changes in solution method for back–of–the–envelope problems. Additionally, subjects and problems were chosen from several different domains.

Methods

Subjects. Two groups of subjects were used, experts and 'intermediates.' In this study, an expert was defined as someone possessing either a doctorate, or a minimum of eight years experience, in his or her field. The four experts included a physicist, a computer scientist, someone with expertise in both physics and computer science, and an expert in a field other than physics and computer science (psychology). An intermediate was defined as a first or second year graduate student. The four groups of intermediates were a computer science student, two computer science students working together, a psychology student, and two psychology students working together.

It was felt that a group of experts and intermediates would provide an interesting body of data on back-of-the-envelope problems to analyze. Beginning graduate students are in a unique position on the continuum between novice and expert. They have more training than a typical undergraduate 'novice,' but have not yet accumulated enough experience, training, and knowledge to be considered an expert. But clearly, they are on their way to becoming experts. They possess the reasoning potential (with only a few exceptions) to become experts in their

field. However, intermediate/expert comparisons are not emphasized in this paper. The discussion will focus on the performance across intermediate subjects.

Procedure. The experts were given a total of six problems. Two of these problems involved physics, two involved computer science, and the remaining two, labeled 'no domain,' tapped knowledge from neither of these domains. The problems used are given in Table 1. The intermediate groups were given four problems: the 'no–domain' problems, the 'bicycle courier' computer science problem and the 'pigeon' psychology problem. All subjects were given pencil and paper for their calculations and asked to give verbal protocols as they were solving the problems. In addition, the experts were audio recorded and the intermediates were both audio and video recorded.

Table 1.

PROBLEM LIST

Computer Science:

At what distances can a courier on a bicycle with a reel of magnetic tape be a more rapid carrier of information than a 56-kilobaud telephone line? Than a 1200-baud line? What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your answer?

Which has the most computational oomph: a second of supercomputer time, a minute of midicomputer time, an hour of microcomputer time or a day of BASIC on a personal computer? How much faith do you have in your answer?

No-Domain:

How much water flows out of the Mississippi River in a day? What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your answer?

How many leaves fall in North America every autumn? What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your answer?

Physics:

About how high does the temperature rise inside a tennis ball when it is hit in a fast serve? What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your answer?

Fueled only by a 2-ounce chocolate bar, how high can you climb is you can turn it into muscular work with 20% efficiency? What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your answer?

Psychology:

A pigeon in a psychology experiment is being presented with a series of geometric shapes on a computer screen. The possible shapes are a circle, square, triangle, and a pentagon. One of these shapes is designated as the correct target shape. For each trial, the pigeon must decide if the shape presented is the target shape. How long would it take a pigeon to peck a button indicating a positive trial (i.e., that the target shape has been presented on the screen?) What is a reasonable upper estimate? A reasonable lower estimate? How much faith do you have in your answer?

Results and Discussion. A characterization of mathematical problem solving developed by Schoenfeld (1985) provides a useful framework for discussing some aspects of the protocols obtained for back-of-the-envelope problems. A summary of Schoenfeld's scheme is provided in Table 2. For each topic -- resources, heuristics, control, and belief systems -- several examples have been chosen to illustrate the role being played in back-of-the-envelope problems.

Table 2.

Knowledge and Behavior Necessary for an Adequate Characterization of Mathematical Problem-Solving Performance

Resources: Mathematical knowledge possessed by the individual that can be brought to bear on the problem at hand

- Intuitions and informal knowledge regarding the domain
- Facts
- Algorithmic procedures
- "Routine" nonalgorithmic procedures
- Understandings (propositional knowledge) about the agreed-upon rules for working in the domain

Heuristics: Strategies and techniques for making progress on unfamiliar or nonstandard problems; rules of thumb for effective problem solving, including

- Drawing figures; introducing suitable notation
- Exploiting related problems
- Reformulating problems; working backwards
- Testing and verification procedures

Control: Global decisions regarding the selection and implementation of resources and strategies

Planning
Monitoring and assessment
Decision making
Conscious metacognitive acts

Belief Systems: One's "mathematical world view," the set of (not necessarily conscious) determinants of an individual's behavior

About self
About the environment
About the topic
About mathematics

Resources. The most obvious factor which will produce differences in subject performance are the resources that a subject brings to a task. This is illustrated by the solution methods of the subjects on the courier problem. The behavior of the two computer science students working together (M.B. and D.W.) resembled that of an expert. They possessed all of the necessary pieces of information to calculate an answer to the problem. The computer science facts (the meaning of baud rate, length of a standard tape, amount of information stored on a tape, etc.) were easily recalled. M.B. immediately states "Let's say a reasonable 1600 bits per inch [density of tape]. And then 2400 [tape length]." Hence, the only estimation required was the speed of a bicycle courier. The behavior of M.B. and D.W. was quite similar to that of S.L., the computer science expert, and both groups of subjects produced realistic answers (S.L.: 38 miles, M.B. and D.W.: 30-60 miles).

In contrast, neither group of psychology intermediates possessed the relevant facts for solution of the problem. S.A. spent very little time working on this problem (See Appendix A for a protocol listening). She does not seem to worry about the facts which would be necessary to calculate an answer, but gives an answer based purely on her (mis)perception of the speed of computers. She simply states "a courier couldn't be faster." It is unclear from the protocol whether her naive conception of computer speed is so strong that it suppresses any

computational effort (the answer seems so obvious that calculations appear completely unnecessary), or if she realizes her lack of knowledge is so severe that her best performance will be a guess based solely on intuition.

The group of two intermediate psychology students working together (B.R. and P.B.), despite a much longer problem solving effort, arrive at the same conclusion. They, however, outline the procedure required to calculate the answer. They are unsure if baud rate means 'bits per second' but decide to operate on that assumption. They then wander through a discussion of many irrelevant facts: speed of electron flow, speed of sound, length of time for a computer to read information, time to print a file, and time to write to a floppy. It seems as though lack of familiarity with the domain prevented the relevant details from appearing immediately salient. B.R. and P.B. do eventually realize that they need to know how much information is contained on a tape, but they are absolutely certain, without estimation or calculation, that it is less than 56,000 bits. (A standard tape can hold approximately $72 * 10^7$ bits.) They state that if this were true then the courier would have less than a second to deliver the information.

Similar to S.A., the most salient detail about computers for B.R. and P.B. seems to be the perception of computers as "infinitely fast." P.B. comments on the speed of computers, "It's almost instantaneous - it's not a perceivable amount of time," and later he states that it takes ten minutes to print a file "but if you write to a floppy -- it's there."

To summarize, B.R. and P.B. realize the pieces of information that are required to calculate a solution, but grossly underestimate the amount of information on a tape and overestimate the speed of computers. S.A. seems governed by her misconception of the speed of computers, and it is unclear whether she understood the necessary procedure. The

computer science students, M.B. and D.W., performed in an expert manner by recalling the necessary facts.

Possession of domain specific knowledge is obviously a critical factor in obtaining a reasonable answer to the courier problem, as with many other types of problems. The protocol of B.R. and P.B. is suggestive of the conclusion that ability to reason in an unfamiliar domain is not entirely hampered by a lack of 'stored facts.' They are able to outline the procedure necessary for a solution of the courier problem. Their main difficulty lies not in understanding or representing the problem situations, but rather in quantizing the problem representation.

The protocol of B.R. and P.B. also suggests that one of the effects of familiarity in a domain is suppression of irrelevant details and foregrounding of relevant facts. While they are able to formulate an appropriate procedure for the courier problem there is much more 'noise' in their protocol than in that of computer science students. B.R. and P.B. consider much unnecessary information about computers before focusing on the relevant pieces of information. It seems that rather than lacking the required knowledge for a solution to the problem the psychology students have trouble identifying the necessary information.

Heuristics. There were several tactics the subjects used to quantize the parameters for back-of-the-envelope problems. Those which have been identified can be characterized as shown in Table 3.

Table 3

Heuristic Tactics

1. Unsure recall of fact, followed by an adjustment.
   Example: A.D.S. on chocolate problem
   "So, and there's 28 grams an ounce, or some such, 24, I don't know. Let me take 25, it doesn't matter much."

2. Unjustified guess.
   Example: J.R. on Mississippi problem
   "Let's say it's a mile wide. And it's probably 50 feet deep."

3. Guess based on experience.
   Example: A.D.S. on Mississippi problem
   "I drove across not too long ago. . . . I don't know any other handle right off the top of my head than just my remembrance of how big the river is."

4. Analogy, usually based on experience.
   Example: M.B. on Mississippi problem
   Comments that he grew up near the Delaware river and that the Mississippi River is wider than the Delaware.

5. Imagery.
   Example: S.A. on Mississippi problem
   "I'm getting confused. I'll just picture it in my mind."

6. Decomposition.
   Example: D.S. on leaves problem
   First estimates the size of a leaf and then calculates increasingly large quantities.

Experiential analogy was a strategy used by many subjects. The object which needs to be assigned a magnitude is compared to a similar object with which one is familiar. When B.R. and P.B. (psychology intermediate) are trying to decide upon a depth for the Mississippi, they call to mind the Long Island Sound and the Cape Cod Canal. These are objects with which they had numbers associated. For example, they recalled that the Long Island Sound is 150 feet deep. (Obviously, this strategy does not always provide the correct answer.) When deciding on the width, B.R. and P.B. compare the Mississippi River to the Oakland Bay Bridge and the Trans Bay Tube. The subjects then compare the object for which they need a number to the object for which they already know a magnitude. Adjustments are then made for perceived differences.

Another strategy which appears very important to all of the subjects is creating an image of the object in their minds before they assign a number to it. S.A. (psychology intermediate) comments while working the Mississippi problem, "I'm getting confused. I'll just

picture it in my mind." Additionally, a picture (an external image) is often drawn of the object. Many of the subjects drew a map of the United States for the leaves problem, and the mouth of the Mississippi River for the Mississippi problem. The map of the U.S. was used to mark the necessary parameters for calculation. Often after drawing the map, the subjects would realize that they originally had forgotten to include Canada and Mexico, interpreting North America as United States. The map would sometimes be used to mark off those areas considered to be heavily forested, thinly forested, or without trees. The drawing of the mouth of the Mississippi River was used in a similar manner. After finishing drawings, comments such as "Okay. Now what do I need to know?" were made. The parameters would then be marked on the drawing.

In order to be able to create an image of an object, subjects decompose the initial quantity in the problem to conceivable objects to which they can then assign values. For example, the number of leaves that fall in North America was always reduced to the number of leaves on a tree. Some subjects even started with the size of a leaf and calculated from that estimate. One of the reasons for this strategy is that a tree or a leaf is easily envisioned by a subject, and hence easily assigned a value. (An equally probable reason is that the answer must be broken up into its component parts in order to be calculated.) There is an interaction between decomposing a quantity into several smaller quantities and visualizing those quantities. Quantities are reduced into component parts until those parts can be assigned values.

A default strategy exhibited by some of the subjects was to make an unjustified guess at some number. This strategy was used in two situations: either the subject had no way of better approximating the quantity, or they felt that it was not necessary to make a more finely honed estimate.

All subjects used some subset of the methods described above, and all subjects envisioned objects. These strategies seem to be a necessary part of solving back-of-the-envelope problems, and to apply across domain.

Control. Two types of control were evident in the solutions to the back-of-the-envelope problems. One of these is a strategy decision; a conscious decision to use one solution method instead of another. This is a choice which will affect the course of the problem solving session. The other type of control is a more localized 'reasonableness monitor.' As a quantity is decided upon or calculated it is evaluated to determine if it meets some criterion of 'reasonableness.'

The first type of strategy decision is illustrated by the four expert protocols on the chocolate problem, shown in Table 4. The physics and computer science expert, A.D.S., solved the problem using a straight forward application of a physics formula and a few estimations. The computer science expert, S.L., attempts to use the same procedure but does not know the necessary energy conversions. He struggles for more than ten minutes on the conversion, never reaching a solution, until he is told to go on to the next problem. The physics expert, D.S., also quickly realizes the applicability of this type of physics approach. He also knows that he does not have the necessary energy conversion. He then backs off from the formula method and instead uses a series of estimations. The estimations require no knowledge of formal physics, but he arrives at a solution remarkably close to that of A.D.S. (I would conjecture that although this solution required no formal physics knowledge, a physics-illiterate person would not arrive at such a reasonable answer using the same method. I believe D.S.'s physics knowledge fed into the accuracy of the estimations.) The psychologist, S.R., realizes immediately that he does not possess the needed physics knowledge and also decides on an estimation procedure. This procedure is more simplistic than that of D.S., but viable nonetheless. The error in this procedure is the result of the grossness of the estimate of

how high one can climb in a day. If that quantity had been broken down into a series of estimations, it is possible that S.R. might have reached an answer closer to that of D.S. and A.D.S.

Table 4

METHODS USED BY FOUR EXPERTS ON CHOCOLATE BAR PROBLEM

ADS (Physics and C.S. expert): Use physics formula (successfully).

2 oz. chocolate = 650 calories = 160,000 joules
160,000 * 0.2 = 560,000
MGH = 546,000
10 * 100 * H = 546,000
H = 546 meters = 1,600 feet

SL (C.S. expert): Use physics formula (unsuccessfully).

2 oz. chocolate = 1000 kilocalories = 1000 joules
tries to use MGH = 1000 and never reaches a 'reasonable' solution

DS (Physics expert): Estimation procedure.

1) Realizes the applicability of formula: potential energy = MGH.
Also realizes he does not know any method for converting 2 oz. of chocolate into energy.

2) Instead makes a series of estimations:
a. Can stay alive for one day on 4 chocolate bars => one bar will keep a person alive for 6 hours

b. 5 times more energy is expended walking than just staying alive => one chocolate bar will keep a person walking for 1 1/5 hours

c. A person can walk 3 k.p.h.

d. Can climb 1/6 as fast as can walk => Can climb 1/2 k.p.h.

Can climb for just over an hour on one chocolate bar, therefore can climb 1/2 km. (= 1650 feet)

SR (Psychology expert): Estimation procedure.

> Determine what percent of daily consumption is represented by one chocolate bar, and how high one can climb in a day.
>
> 2 oz. chocolate = 300 calories
> daily calories = 4000 => chocolate bar = 5% daily calories
> can climb 4000 feet in one day => can climb 200 feet fueled by one chocolate bar

D.S. makes a control decision which allows him to reach a reasonable answer. Because he is consciously monitoring his solution process, he realizes the inadequacy of his knowledge for the formal physics approach. D.S. then backs off from attempting to apply a physics formula and tries another approach. This second approach utilizes less formal knowledge and estimations rather than formulas. D.S. is hence able to tap a different set of resources and successfully reach a solution. In contrast, S.L. never stops to evaluate his progress on the problem and simply runs in place for most of the session. These protocols provide an interesting example of the diversity of solution methods obtained for back-of-the-envelope problems.

A more omnipresent and localized type of control was the constant monitoring for the reasonableness of a quantity either chosen or calculated. When a quantity was formed from several lesser quantities, it was often again held up against a new standard for reasonableness. Frequent comments as numbers were being generated included "Is that reasonable?", "Does that make sense?", "I don't like that, it seems very unreasonable", and "Let's think about this for a minute, it is reasonable?"

When the answer to a problem was reached it was often inspected for reasonableness by comparing it to another known quantity. M.B. and D.W. (computer science intermediates) had reached an answer of 10^{13} for the leaves problem.

D.W.: Do you have any idea how large this number is?
M.B.: Is it bigger than the national debt?
D.W.: Yes.
M.B.: Then it's probably inaccurate. . . . Is there one real tree for every person [in the United States]?

In this excerpt, they have compared their answer both to the national debt and to the population of the United States in an attempt to get a handle on the magnitude of their answer.

A.D.S. (physics and computer science expert) after calculating that there were 50 billion trees in the United States says,

> "People in the U.S. Imagine each person as a tree. I'm really thinking about my experiences growing up in Colorado. Trying to attach a person to a tree. Let's see how that goes. I don't think that's going to get me anywhere. 50 billion. That's reasonable enough to go with I guess."
>
> Later, when judging confidence in his answer, he states:
>
> "I said how many trees did I have, 50 billion trees?... I'm looking for independent estimates of these guys. What's the gestalt of 50 billion trees?"

A.D.S., as M.B. and D.W., is attempting to inspect the answer he has reached by comparing it to other quantities which he already knows. Because the magnitude of the answer is so large that it is hard to comprehend, the subjects seem compelled to compare the answer to another quantity for which they have some associated meaning.

Similarly, B.R. and P.B. (psychology intermediates) when judging confidence in their answer to the leaves problem, also try to 'grok' 10^{14}. B.R. remarks, "I can't conceive of numbers that high. I've never counted that high [laughs]. I've never had that many of anything." The same subjects solve the Mississippi problem after having worked out an answer for the leaves problem. B.R. suggests a confidence rating of 10%.

P.B.: You're 60% sure of the leaves and only 10% confident of this? At least we're dealing with numbers we can comprehend.
B.R.: You can comprehend a cubic mile? I'm skeptical.

These examples illustrate the need subjects have to judge their answers for reasonableness. It seems as though inability to envision the magnitude of their answer often drives the subjects to compare their answer to other quantities

Belief systems. One of the reasons for choosing the two different areas of graduate students was their different positions along the continuum of quantitative/non–quantitative disciplines. Computer science is considered a mathematical field, whereas psychology is often thought of an area in which mathematical ability is not essential. Many psychology students are not only disinterested in mathematics (or their perception of what mathematics entails), but are also math phobic or at least math shy. Computer science students, on the other hand, seem to be comfortable with the quantitative aspects of their work. In fact, it may be this quantitative aspect which originally attracts many students to computer science.

To some degree, these attitudes are reflected in performance on back-of-the-envelope problems. They are more evident, however, in the subjects' perception of their performance, rather than in actual competency. None of the subjects was unable to perform a necessary computation or realize a plausible approach to the problem. (With the possible exception of S.A. on the courier problem. It is unclear whether she would have figured out the necessary computational steps.) In judging faith in their answer, however, subjects in different domains varied in their confidence rating, despite having performed virtually identical computations.

The pigeon question produced quite brief and similar protocols for all the subjects. (This problem was at first misunderstood by all of the subjects, and none of the subjects actually did a series of estimations. Apparently, it was not a clearly worded question.) However, despite the similarity among the responses, there was a significant difference in the amount of confidence the subjects had in their answers. After S.A., a psychology student, has

realized what the question is asking for (reaction time), her protocol is quite brief (see Appendix B). Although she realizes some of the relevant parameters -- reinforcement, distance from the button, etc. -- and divides the reaction time into at least two stages, recognition and response time, she does not take these factors into account. She simply states "Okay, half a second." S.A. does claim, however, to have more faith in this answer than she does for the leaves question. This is despite the fact that she performed a series of computations for the leaves problem.

B.R. and P.B., also psychology students, throw around some response times they are familiar with such as "100 milliseconds or something" for a neuron to fire to a visual sensation, and "200 milliseconds to execute an eye movement." As S.A., they then simply state that the response time is "less than one second." (I am not suggesting that these number are not reflected in the answer in some way, simply that they did not explicitly estimate X seconds for perception +Y seconds for decision making +Z seconds for motor response=total response time. This, incidentally, was desired behavior.) They set an upper limit of 1/2 second and a lower limit of 100 milliseconds. Their confidence rating in this brief calculation was a high 90%.

M.B. and D.W., computer science students, spend less time on the pigeon question then they did on any of the other questions. D.W. clocks M.B. as he strikes the table "as if I'm a pigeon recognizing something." This takes about 1/2 second and their final answer is in the range from 1/2 second to 1 second. M.B. suggests that they have more faith in this answer than in the leaves problem, and D.W. responds "I don't know anything about pigeon psychology." Their confidence is finally decided upon at 35%.

All of the answers given were quite similar, and none of the subjects spent much time calculating an answer. M.B. and D.W. have only 35% confidence, despite having performed a mini-simulation, while B.R. and P.B. have 90%. The extreme variance in these confidence

ratings suggests that the subjects are responding to their confidence in their ability in a domain, rather than their confidence in the computations they have just performed.

In general, having more confidence in an area which one has spent time studying is a reasonable technique. Given that someone is trained in a particular field, chances are that they will perform better in that area then in others. It only becomes a problem when it acts as a hindrance in areas perceived to be outside one's expertise. Realization of one's shortcomings can be valuable, but can also be unnecessarily restrictive.

Before solving the problems, S.A. (psychology intermediate) was very concerned about her ability to do the necessary mathematics for the back-of-the-envelope problems. However, she exhibits mathematically sound, but unsophisticated, behavior. When calculating the number of trees in a square mile, she states:

> Every 8 feet could have a tree. So divide one mile, find out how many 8's that would be [divides 5,280 by 8]. Is that true? Yea, already. Every one mile you could have 660 trees.
>
> ... So now I'm saying if I space them out across the mile like this and say every 8 feet could have another tree, and how many quadrants [draws a square with 660 marked on each side and divides the square into quadrants]? So that would be 660 again. So 660 squared. So it would be like this, 660 this way and 660 that way, and everyone would have a tree. Already. Maybe I can get a job with the forest service [squares 660].

The statement "find out how many 8's there would be" sounds surprisingly like a school child. She finds out the number of trees in a square mile by drawing a square and dividing it into quadrants. She seems to have returned to the meaning of the concept of squaring, rather than accessing a stored "squaring schema." Despite this lack of mathematical sophistication, S.A. manages to arrive at answers to all of the problems.

It is also notable that S.A. does not discard any digits; she maintains all the numbers in her final answer. She does this in all her protocols, never giving an answer in scientific

notation, as did everyone else. Her answers are not wrong; they are simply less polished than the others.

When S.A. is solving the Mississippi problem, she associates what she is doing with statistics, most likely her only recent quantitative experience. "Already, the river flows one foot per second. This is like statistics." She has figured out the necessary parameters to solve the problem and then states:

> So I've got to figure out a little math problem here. So I'll take something I know that I can figure out. If I can't figure things out, I can always try to reduce it to a more simple way of trying to figure it out.

Despite S.A.'s lack of mathematical sophistication and the fact that she was worried before the protocol session about being able to perform the necessary mathematical computations, she performs competently on all of the problems. She manages quite well to work around all the mathematical obstacles she encounters. However, she probably would not have agreed to give the protocols had she been asked to 'solve some math problems.' S.A. is a good example of a person who should consider herself mathematically untrained rather than mathematically incompetent.

Summary. The four categories of resources, heuristics, control, and belief systems have provided a framework for discussing some of the aspects of the reasoning involved in back-of-the-envelope problems. Resources, of course, dramatically affect the solution processes used by the subjects. Back-of-the-envelope problems may be an area in which lack of domain specific knowledge may be compensated. For example, on the chocolate problem two experts, D.S. and S.L., managed to maneuver around their lack of knowledge by using estimation procedures. This may provide an interesting source of data on reasoning from incomplete knowledge. The types of heuristics the subjects use were divided into several categories. The two most common strategies were to compare the object at hand to some

other object with which one has a value associated, and to create a mental image of the object to be quantized. Two types of control were observed in the subjects. The first was conscious strategy decision to use a particular method for solving a problem. The second was a more localized monitor for the reasonableness of quantities which were being estimated. The belief systems that the subjects have about their competence in different domains affect confidence ratings of their answers. This is true regardless of the actual computations the subjects perform.

A Model of BEP

The previous section examined some of the differences between subjects whenever either the expertise of the subjects or the problem domain varied. It would also be useful to characterize the similarities among subjects on back-of-the-envelope problems. A model will be developed in is section in order to examine the processes and knowledge required to solve this class of ill-structured problems.

The framework provided by FERMI, "Flexible Expert Reasoner with Multi-domain Inferencing," (Larkin, Reif, Carbonell, and Gugliotta, 1985) has been adapted to model the solutions to back-of-the-envelope problems. FERMI stores knowledge and problem solving methods in a hierarchy according to their level of generality. This is an especially useful feature for this type of problem, in addition to the particularly *apropos* name of the system. FERMI was originally designed as an expert system, not as a model of human behavior. In addition to providing a framework for the discussion of the reasoning involved in back-of-the-envelope problems, this analysis will show that FERMI is a viable model of human cognitive activity. There are several further ways in which FERMI will be extended. The first involves the addition of more "everyday" type of knowledge, and more general methods such as estimation. Secondly, FERMI will be shown to provide an adequate model for a domain of problems not previously considered. Thirdly, in addition to modeling human cognitive activity it will be shown that individual subject behavior can be modeled by adjusting the knowledge base available to the system. Finally, the protocols previously discussed support several assumptions in the design of the system; primarily the hierarchical structure of the problem solving methods in the solution to a problem.

FERMI is a computer-implemented expert reasoner in the natural sciences that encodes its declarative and procedural information hierarchically at the appropriate level of generality.

In particular, the system includes two related hierarchies, one of scientific principles and one of problem-solving methods. This should provide hypotheses of the manner in which skilled human experts separate and use knowledge according to its generality.

FERMI has been implemented in the schema-representation language SRL (Fox, 1979; Wright and Fox, 1983). The system uses schemas (Minsky, 1975; Bobrow & Norman, 1975), data structures composed of slots and fillers for storing related knowledge. Any slot in a schema may have associated information about how the slot may be filled, such as default values and constraints. Slots in FERMI may also have associated pullers, i.e., pieces of code to be implemented whenever the system needs to fill a slot about which it has no stored information. Hierarchies are created by connecting schemas with *isa* links which indicate class membership. When a schema A is connected by an *isa* link to a second schema B, then A automatically inherits all the contents from the schema B. The *isa* relation is also transitive. That is, if A *isa* B and B *isa* C, then B inherits directly the contents of C, and A inherits from B both the original contents of B and all the knowledge that B inherited from C. This inheritance allows knowledge common to a variety of schemas to be encoded only once.

FERMI is based on research of how information is structured in the physical sciences (Chi, Feltovich, and Glaser, 1981; Reif and Heller, 1982). Physical scientists can identify general principles and problem-solving methods (e.g., energy principles or decomposition methods) as well as specific instantiations (e.g., decomposition of vectors into components). They can also distinguish between more and less general principles or methods. (For example, the statement "path integrals of scalar-field differences are path independent" is quite general, while the statement "pressure drop in a static fluid is path independent" is specific to the domain of fluid statics.) FERMI's knowledge is thus organized into two distinct schema hierarchies, one encoding scientific principles of different levels of generality, and the other encoding problem–solving methods of different levels of generality.

In the current work, the hierarchies of FERMI will be extended further to include even more general reasoning methods. The principles and methods used by FERMI are more or less *scientifically* general. They do not embody the everyday reasoning skills used by a non-scientist. Consequently, the interaction between the scientific and general knowledge and strategies cannot be modeled. One of the protocols which will be discussed shows a scientist interweaving the two types of knowledge in order to arrive at an answer.

FERMI's general knowledge is stored in general "quantity schemas" and in associated general "method schemas." A general quantity schema contains pointers to one or more general methods. These pointers are inherited by all quantities related to that general quantity by any chain of *isa* links. Likewise, FERMI's domain-specific knowledge is stored in domain-specific quantity schemas and in associated local methods called "pullers." There pullers contain procedural knowledge about how to fill a slot when it is empty, and no inheritable value is available.

If domain-specific knowledge alone fails to solve a problem, FERMI tries more general methods. However, the general methods alone cannot usually solve the problem alone and require specific information. This information is recursively supplied by the domain-specific quantity schemas and their pullers. This creates an interesting interaction between the domain-specific and general knowledge.

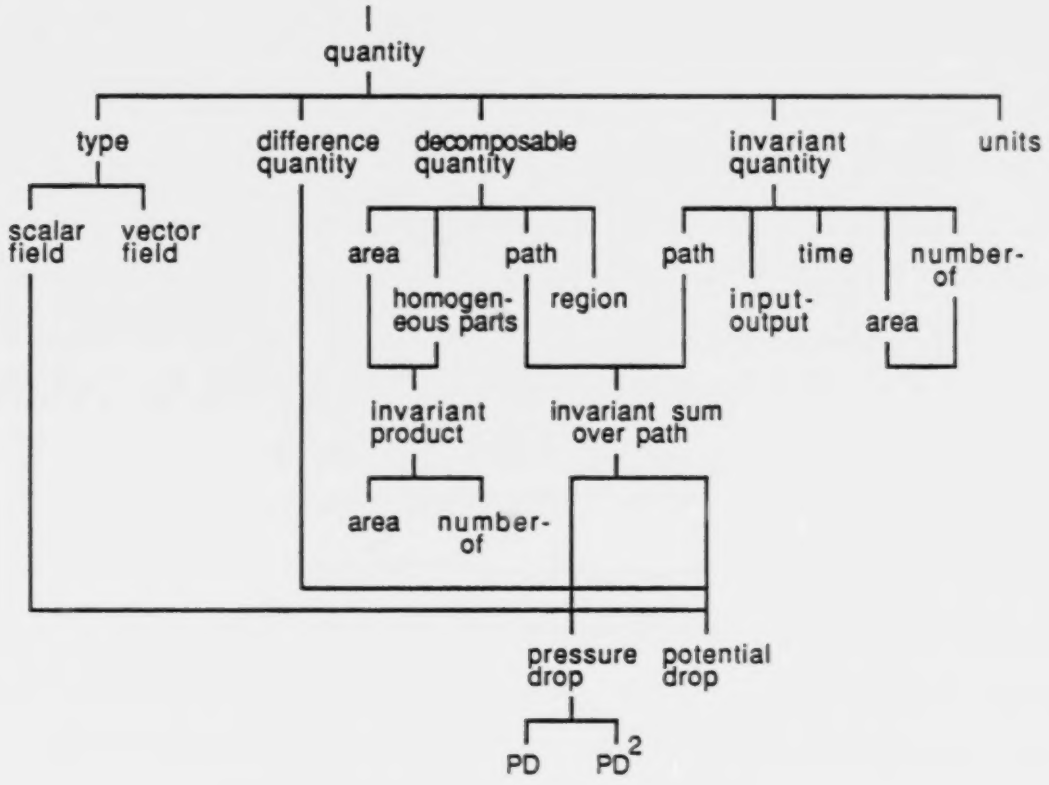

Figure 1. FERMI's hierarchy of quantities.

Figure 1 shows part of FERMI's hierarchy of quantity schemas, part of the more encompassing hierarchy of entity schemas illustrated in Figure 2.

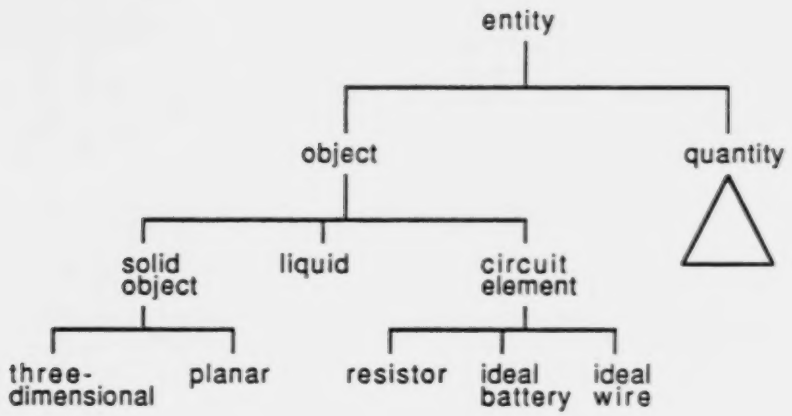

Figure 2. FERMI's hierarchy of entities, including the hierarchy of quantities from Figure 1.

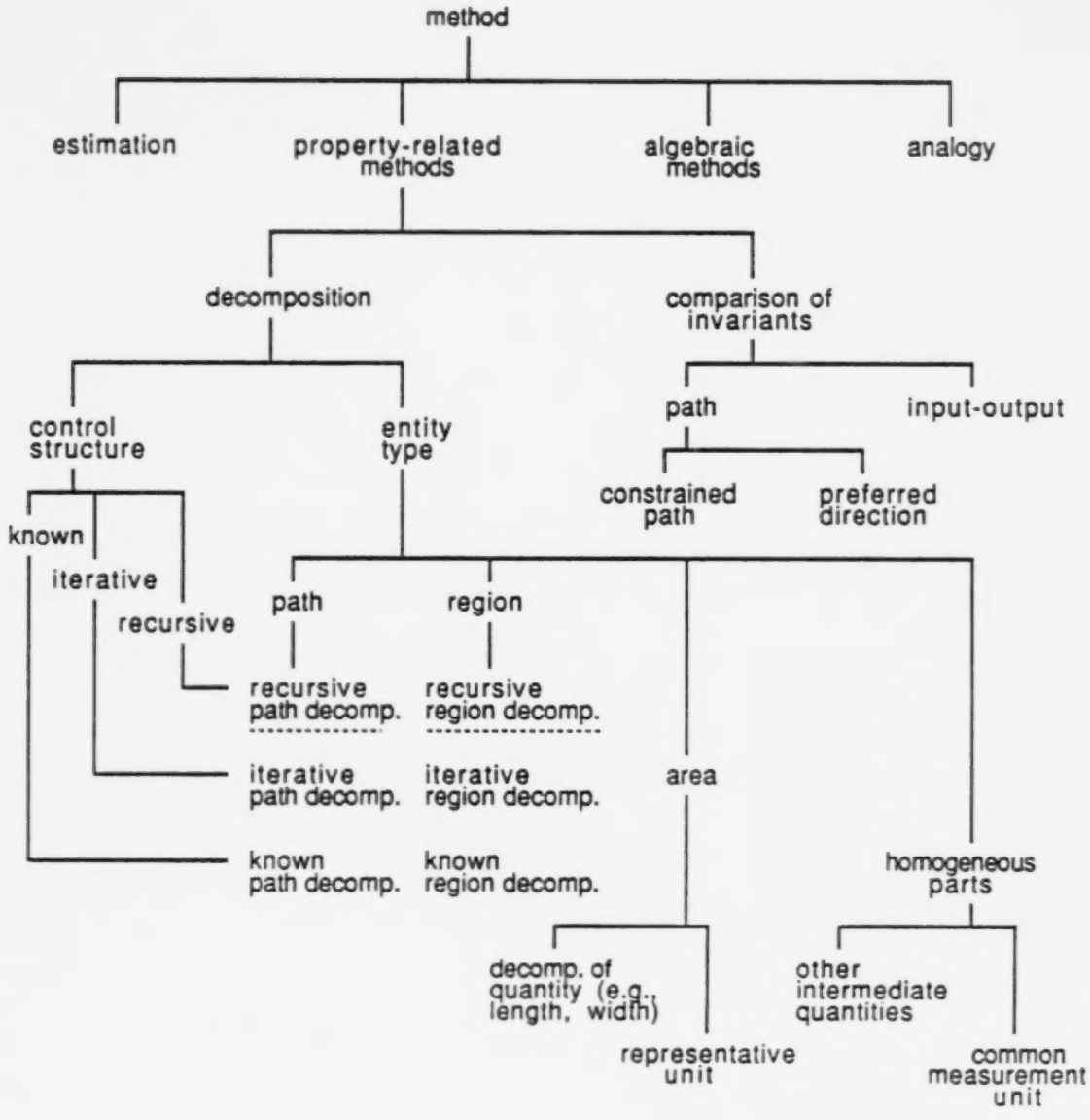

Figure 3. FERMI's hierarchy of major problem-solving methods

Similarly, the hierarchy of method schemas in Figure 3 is only a part of the broader hierarchy of action schemas in Figure 4.

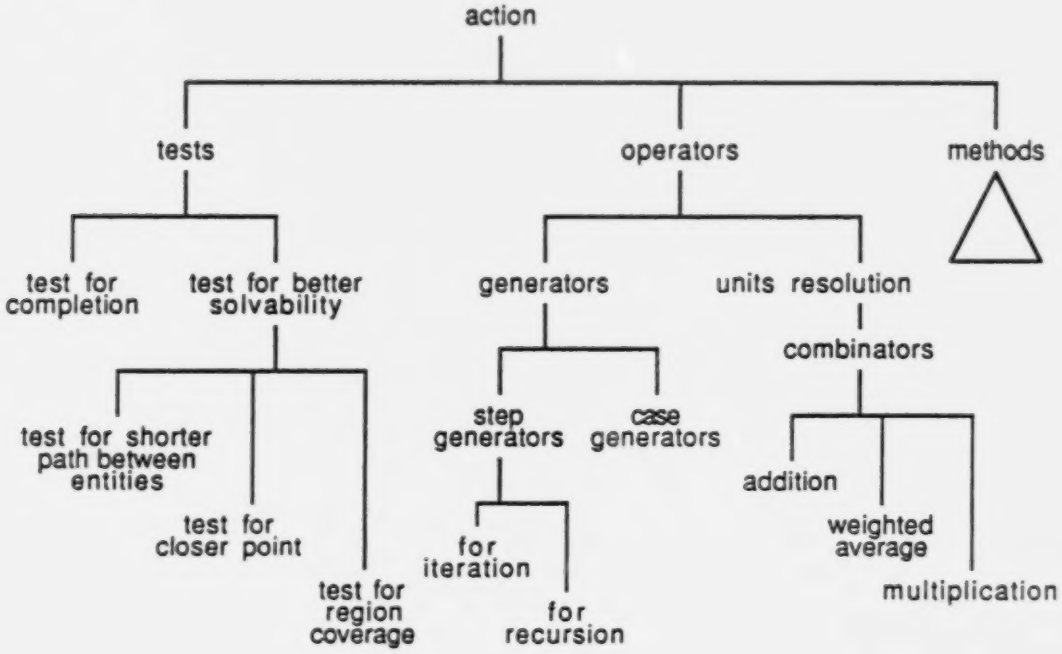

Figure 4. FERMI's hierarchy of actions, including methods included in Figure 3.

Figures 5.1 - 5.5 show a trace of a solution to the problem produced by an extended version of FERMI. The problem is: "How many leaves fall in North America every autumn?" This is an example of a problem requiring only general knowledge in order to calculate an answer.

```
G1:  number-of1 [leaves, N.A.]
          Look-up: empty
     G2:  Apply-pullers
               none
     R2:  falls
          G3:  Apply-methods {homogeneous-parts decomposition, estimation}
                    Apply method: hpd
                    number-of1 = expression1
                         [* number-of2 number-of3]
          R3:  number-of1 = expression1
               G4:  evaluate-expressions-for-no-of1
                         OR {expression1}
                    G5:  evaluate expression1: (* number-of2 number-of3)
                              AND {number-of2 number-of3}
                         G6:  number-of2
```

This part of trace elaborated in Figure 5.2

```
                    R6:  number-of2 = 6,750 leaves /tree

                    G15:  number-of3
```

This part of trace elaborated in Figure 5.3.

```
                R15:  number-of3 = 6.539 * 10^11 trees/N.A.
           R5:  expression1 = 4.393025 * 10^11 [* 6,750 6.539 * 10^11]
      R4:  evaluate-expressions-for-number-of1: 4.393025 * 10^11

R1:  number-of1 = 4.393025 * 10^11 leaves/NA
```

Figure 5.1. Trace of FERMI's solution of a problem (main steps)

Figure 5.1 shows the main goals and results, with subsequent figures giving more details. The trace is organized as nested sets of goals and corresponding results. In Figure 5.1, the desired quantity called "number-of1" is found in three steps. First, "look-up" fails because the number is not already available to the system. In correspondence, it is unlikely

that many people have stored and can recall the number of leaves that fall in North America. FERMI then tries to use pullers. As there are none available for this quantity type, this technique also fails. The attempt to use pullers with number-of quantities will be omitted subsequently as it will always fail. As a third step, FERMI identifies applicable methods. The pointer to these methods are inherited by number-of from the general quantity schema "quantity decomposable into homogeneous parts." In this case, applying homogeneous parts decomposition (hpd) to number-of1 produces expression1, (number-of2 number-of3), where number-of2 and number-of3 are, respectively, the number of leaves on a tree and the number of trees in North America. FERMI has thus decomposed the initial quantity into two lesser component quantities.

In order for the method hpd to apply, it must be able to decompose the initial number–of quantity into two component number-of's. First, a relationship must be found between the two original objects and an intermediate object. This relationship must allow for a number-of link to be created. For example, in the current trace, FERMI finds that leaf and tree are connected by the relationship "grows-on," or conversely "grows," therefore the number of leaves on a tree can be calculated. FERMI must then find a relationship between tree and North America. There is no strong, direct link as for leaf and tree; however, a common unit of measurement can be found using the concept of area. FERMI has thus decomposed the quantity of leaves in North America into the smaller quantities of leaves on a tree and trees in North America. One of the computability requirements of this method is that the component quantities must be lesser quantities than the original quantity. Finally, the combination function of this method indicates that the component quantities must be multiplied together.

In (G3, R3), FERMI evaluates the single expression generated, yielding the desired quantity. (G4, R4) requires the *AND* subgoal to find values for both number-of2 and

number–of3. The actual numbers shown in this trace were taken from the protocol of S.A. (psychology intermediate) solving the problem (see Appendix C).

```
G7:  number-of2 [leaves, tree]
       lookup: empty
   G8:  Apply methods {hpd, estimation}
          Apply method: hpd
          number-of2 = expression1
             [* number-of4 number-of5]
   R8:  number-of2 = expression2
      G9:  evaluate-expression2:
             OR {expression 2}
         G10:  evaluate expression2: (*number-of4 number-of5)
                AND {number-of4 number-of5}
            G11:  number-of4 [leaves,branch]
                  Lookup: empty
               G12:  Apply methods {hpd, estimation}
                     Apply method: hpd
                     fails
                     Apply method: estimation
                     number-of4 = 750
               R12:  number-of4 = 750
            R11:  number-of4 = 750
               G13:  number-of5 [branches, tree]
                     Lookup: empty
                  G14:  Apply methods {hpd, estimation}
                        Apply method: hpd
                        fails
                        Apply method: estimation
                        number-of5 = 9
                  R14:  number-of5 =9
               R13:  number-of5 = 9
         R10:  expression2 = 6,750 [* 750 9]
      R9:  evaluate-expressions-for-number-of2:  6,750

R7:  number-of2 = 6,750
```

Figure 5.2. Trace of FERMI finding number-of2.

In Figure 5.2, the process of finding the number of leaves on a tree is shown. Again, look-up does not supply an answer; consequently, the method hpd is applied. Number-of2 is decomposed in a similar manner to number-of1, with expression2 resulting. Number–of4 and number-of5 are respectively the number of leaves on a branch and the number of branches on a tree. The decomposition is slightly simpler than the first in that "branches" is directly

connected to both "leaves" and "trees." Note that this decomposition provides the possibility of simulating individual differences. If the branch links were omitted from the knowledge structure available to FERMI, then the decomposition could not occur. Apparently, the number of leaves on a tree is not always considered a decomposable quantity as some subjects omitted this step.

When trying to find number-of4, FERMI again tries to apply hpd. However, this method fails because there is no intervening object in the entity hierarchy between leaf and branch. The method of estimation is therefore applied to this quantity. Estimation will be treated in this paper as a black-box procedure. It will simply provide a number when appropriately applied. Estimation could conceivably be used to generate an answer for any desired quantity; however, its use is generally constrained by at least two factors. First, estimation is used more often for quantities which are not easily or possibly measured. For instance, in the current problem, the number of leaves on a tree is not a number simply to count or measure. Furthermore, even if one did manage to count the leaves on a given tree, this would not indicate that this is a reasonable number to represent the average number of leaves on the average tree. In this case, estimation seems as a viable a method as measurement. On the other hand, in a physics problem, this is not often a good approach because the problems usually deal with specific physical situations. When this is true, there is a precise quantity needed which can be calculated or measured.

Number-of5, the number of branches on a tree, is found in a similar manner of number–of4.

G15: number-of3 [trees, NA]
Lookup: empty
G16: Apply methods {hpd, estimation}
number-of3 = expression3
[/ area1 area2]
R16: number-of3 = expression3
G17: evaluate-expressions-for-number-of3
OR {expression3}
G18: evaluate expression3: (/ area1 area2)
AND {area1 area2}
G19: area1 [NA]

This part of trace elaborated in Figure 5.4

R19: area1 = 1.5 * 106 sq. miles

G27: area2 [tree]

This part of trace elaborated inFigure 5.5.

R27: area2 = 64 sq. feet

R18: expression3 = 6.539 * 10^{11} [/ 1.5 * 10^{6} sq. miles 64 sq. feet]
R17: evaluate-expressions-for-no-of3: 6.539 * 1011

R15: number-of3 = 6.539 * 10^{11} trees/NA

Figure 5.3. FERMI's trace of finding number-of3.

In Figure 5.3, the number of trees in North America is calculated. Once again, the method hpd is applied to the desired quantity. In this case, trees and North America are not directly connected, or connected via an intermediate object; therefore, the concept of area is used to connect the two objects. The number of trees in North America is decomposed into the area of North America, which must be divided by the area of a tree. The subgoal is then set to find the area of North America. This is shown in Figure 5.4. The method which is applicable in this case is called "area decomposition." This method decomposes the area of North America into its length and width, which must then be multiplied together. The length and width of North America are estimated similarly to the number of leaves on a branch and the branches on a tree.

```
G19:  area1 [NA]
      lookup: empty
   G20:  Apply methods {area decomposition, estimation}
         Apply method: ad
         area1 = expression4
            [* length1 width1]
   R20:  area1 = expression4
      G21:   evaluate-expressions-for-area1
            OR {expression4}
         G22:  evaluate expression4:  (* length1 width1)
               AND {length1 width1}
            G23:  length1
                  Lookup:  empty
               G24:  Apply methods {hpd, estimation}
                     Apply method:  hpd
                     fails
                     Apply method:  estimation
                     length1 = X
               R24: length1 = X
            R23: length1 = X
               G25: width1
                     Lookup: empty
                  G26: Apply methods {hpd, estimation}
                        Apply method: hpd
                        fails
                        Apply method: estimation
                        width1 = X
                  R26: width1 = X
               R25: width1 = X

      R22:  expression4 = 1.5 * 10^6 sq. miles [ * x miles x miles]
   R21:  evaluate-expressions-for-number-of4:  1.5  * 10^6 sq. miles

R19:  area1 = 1.5 * 106 sq. miles
```

Figure 5.4. FERMI's trace of finding area1.

Units of measurement become important when multiplying two quantities. Multiplying a quantity of miles by another quantity of miles must result in an answer involving square miles. Additionally, if two numbers are estimated in different units, one of the quantities must be converted in order for the mathematical operation to be performed. Area decomposition therefore must pass both its arguments and their units of measurement to a type of action called an operator. This operator, called "units resolution," takes as input two quantities with their associated units and a mathematical operator. It then produces the correct quantity and

units indicator. This step of resolution is not reproduced in the trace, but occurs any time a mathematical operation is performed.

```
G27:  area2 [tree]
      lookup; empty
   G28:  Apply methods {area decomposition, estimation}
         Apply method: ad
         area2 = expression5
            [* length2 width2]
   R28:  area2 = expression5
   G30:   evaluate-expressions-for-area2
         OR {expression4}
      G31:  evaluate expression5: (* length2 width2)
            AND {length2 width2}
         G32:  length2
               Lookup: empty
            G33:  Apply methods {hpd, estimation}
                  Apply method: hpd
                  fails
                  Apply method: estimation
                  length2 = 8 feet
            R33:  length2 = 8 feet
         R32:  length2 = 8 feet
         G34:  width2
               Lookup: empty
            G35:  Apply methods {hpd, estimation}
                  Apply method: hpd
                  fails
                  Apply method: estimation
                  width2 = 8 feet
            R35:  width2 = 8 feet
            R34:  width2 = 8 feet
      R31:  expression5 = 64 sq. feet [* 8 feet 8 feet]
   R30:   evaluate-expressions-for-number-of5: 64 sq. feet
R27:  area2 = 1.5 * 10^6 sq. miles
```

Figure 5.5. FERMI's trace in finding area2.

In Figure 5.5, the area of a tree is found in a manner similar to the area of North America. In order to calculate R17 in Figure 5.3, the number of trees in North America, the area of North America is divided by the area of a tree. Note that the resolution of units is also critical to this operation. This result is in turn combined in Figure 5.1 with the number of leaves on a tree to produce the final answer, the number of leaves on a tree.

The leaves problem demonstrates that problems of some complexity can be solved solely with general methods. These problems, however, require extensive use of procedures such as estimation, about which we know very little. The protocols of experts and intermediates do not seem to differ significantly on these problems. This supports the conclusion that these general methods are accessible to both groups of subjects, and that the organization of the knowledge structure required for solution of this type of problem is similar.

Without providing a detailed trace a solution of D.S. (physics expert) to a second problem will be discussed. See Appendix D for a complete listing of the protocol. The problem is as follows: "Fueled only by a two-ounce chocolate bar, how high can you climb if you can turn it into muscular work with 20% efficiency?"

In Episode (1) of the protocol, the expert outlines the method he would like to use to solve the problem, basically using the formula for potential energy. This corresponds to the use of domain-specific pullers in the FERMI system. These pullers, however, would fail because the expert does not know the energy content of a chocolate bar and knows no other way of getting this necessary quantity. He then resorts to the general method of a series of estimations. It is conjectured that his solution does not proceed exactly as it would if he were without physics knowledge, as he still has access to and uses domain-specific pieces of knowledge. He states in Episode (3) that he knows it takes 100 joules per second to stay alive and then uses this number for comparison in Episode (5). Despite the fact that this subject is using general method, he is still utilizing domain-specific knowledge.

This ability to access domain-specific knowledge while using general methods seems to be an aspect of expertise. In addition, an expert's knowledge may initially guide the choice of the general method. While intermediate or novice protocols were not collected for the

chocolate problem (or other physics problems), it would be interesting to make comparisons in order to test these conjectures.

It was noted on the leaves problem (and other 'non-technical' problems) that expert and intermediate solutions do not seem to differ significantly. However, it was conjectured that for domain-related problems that expert/intermediate/novice differences would appear even when general methods were being used. These differences would result from the differing degree to which subjects could access domain-specific knowledge. This suggests further research into the contribution of knowledge to general quantitative reasoning tasks. The question is whether experts, intermediates, and novices differ in their approaches to solving problems when their respective knowledge is inadequate. It is hypothesized that the answer to this question is "yes," for several reasons. First, an expert's aborted attempt at a domain-specific method may indicate a viable general method through the hierarchical organization of the methods. The intermediate or novice problem solvers may not have a pointer to the knowledge structure in this way. Secondly, even while using general methods experts still have access to domain-specific pieces of knowledge not available to the novice or intermediate. Finally, knowledge reflecting differing degrees of generality may be organized differently across levels of expertise.

References

Bentley, J. L. (1984). The back of the envelope. *Communications of the ACM, 27,* 180-184.

Bentley, J. L. (1986). The envelope is back. *Communications of the ACM, 29,* 176-182.

Bobrow, D. G., & Norman, D. A. (1975). Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 131-150). New York: Academic Press.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Fox, M. S. (1979). On inheritance in knowledge representation. *IJCAI VI Proceedings* IJCAI-79.

Greeno, J. G. (1978). A study of problem solving. In R. Glaser (Ed.), *Advances in instructional psychology,* (vol. 1). Hillsdale NJ: Lawrence Erlbaum Associates.

Greeno, J. G., & Simon, H. A. (in press). Problem solving and reasoning. In R. C. Atkinson, R. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Steven's handbook of psychology* (revised edition). New York NY: John Wiley & Sons.

Larkin, J. G., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science, 4,* 317-345.

Larkin, J. H., Reif, F., Carbonell, J., & Gugliotta, A. (1985). *FERMI: A flexible expert reasoner with multi-domain inferencing.* Pittsburgh, PA: Carnegie-Mellon University Department of Psychology.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision.* New York, NY: McGraw-Hill.

Morrison, P. (1963). Fermi questions. *American Journal of Physics, 31,* 626-627.

Purcell, E. M. (1983). The back of the envelope. *American Journal of Physics, 51,* 11.

Reif, F., & Heller, J. (1982). Knowledge structure and problem solving in physics. *Educational Psychologist, 17,* 102-127.

Reitman, W. R. (1965). *Cognition and thought.* New York NY: John Wiley & Sons.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* New York, NY: Academic Press.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4,* 181-202.

Voss, J. F., Greeno, T. R., Post, T. A., & Penner, B. C. (1983). Problem solving skill in the social sciences. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (vol. 17, pp. 165-213). New York, NY: Academic Press.

Voss, J. F., Tyler, S. W., & Yengo, L. A. (in press). Individual differences in the solving of social science problems. In R. F. Dillon & R. R. Schneck (Eds.), *Individual differences in cognition.* New York: Academic Press.

Wright, K., & Fox, M. (1983). *The SRL User's Manual* . Pittsburgh, PA: Carnegie-Mellon University Robotics Institute.

Appendix A.

S.A. - COURIER PROBLEM:

At what distances,
well I think that a courier can't go very fast on a bicycle so the only real distance that the courier could be faster on a bike would be if he didn't ride the bike.
If he just sat there and went like that [hits table].

But then you might, I don't know if you have to take the time to put the reel of magnetic tape on the, on whatever the machinery is, to get the information, if we're just passing the information from one place to another or if we're being able to look at it at the same time.

This telephone line looks like it would probably get the information in a cluster and be able to look at it, or hear it, pretty much simultaneously compared to getting a reel of information and having to mount it onto some sort of device to then have access to it.

So I would say a courier couldn't be faster.
But it says carrier it doesn't say interpret it.
A lot of these things I think don't matter.

Well, pretty much faith in that answer.
I hope that I don't get kicked out of graduate school for this.

Appendix B.

S.A. - PIGEON PROBLEM:

I see what you're saying now, how fast does the pigeon recognize that the shape is the target, and then how fast does the pigeon respond.
Okay.

Well, I guess it depends on how close to the lever, the pecking button, the pigeon is.
And what the reinforcement is.
So it would depend on some things.
[writes down] Distance from the button,
I might be missing this completely.
What reinforcements has gotten in past.
And how long the stimulus on screen.

So I estimate the pigeon will peck on the button...
Well I know these pigeon are very fast.
Okay, a half second.
And then, that's an upper estimate.
No, that's not an upper estimate. Average.
But I don't think the upper and lower limits are very different.
Once the pigeon has learned.
Oh, might be...

How much faith... Okay.

Appendix C.

S.A. - LEAVES PROBLEM:

Well, the first thing that comes to my mind is that I would want to see a map of the tree population on North America and how dense the trees are in certain areas to be able to get a square mile estimate of trees.
And then I would want to know how many trees are in each square mile.
So, I guess, I don't know what I should be writing down on this piece of paper.
And I also know that some leaves don't lose their, some trees don't lose their leaves, and other trees do.

So I would say that, I don't know how many squre miles there are in America.
But I would want to look at a map and say...
[draws an X and Y axis and then draws the shape of Michigan around those lines]

This is a map of Michigan because that's where I was born.
200 miles times 400. 80,000 square miles.
So I'm trying to figure out how many square miles there are in Michigan.
And then I would want to know, that's an average state.

[multiplies 80,000 by 50]
Oh boy, so there's 4 million square miles in the United States.
But North America, that includes Canada too.
But a lot of that is above the tree line.

Now, so we'll double that and say there's 8 million [square miles in North America].
And then how much of it has trees.
This is a map of the United States [draws a map of the United States].
This is Canada.
Tree line, probably goes like that.
Not a lot of trees over there [draws tree lines in Canada and Mexico].

So then I would say where do I think the biggest conglomerations of trees that the leaves fall are there.
To achieve the number of square miles with trees.
Okay, I'd say about one third [visually she has marked off the top and bottom thirds of the map of North America].

So one third of 8 million is 2,500,000 square miles.
So that's too much, because there's lakes and roads and cities and buildings there too.
So I'd probably lower it down to 2 million miles, square miles.

But then there's mountains where there's not a lot and...
There's a high altitude that's above the treeline so let's see [reduces to 1.5 million sq. miles].
Then I can figure out how many trees in a square mile.

Well the forest around the house that I lived in is a tree land. [inaudible.]
If I know how many trees are in a square mile I'd fit... How many trees would fit in a square mile?
One mile has 5,280 feet [writes 5,280 ft. = 1 m.].
Let's see how many miles, I'm trying to think.
If a tree was about 2 feet wide.
And most trees are about 6 feet wide [writes down 6 feet apart].
Then if I had one mile that would be 8 feet.
Every 8 feet could have a tree.
So divide one mile, find out how many 8's that would [divides 5,280 by 8].
Is that true? Yeah, alright.
Every one mile you could have 660 trees. About 6 feet apart.
So now I'm saying if I space them out across the mile like this and say every 8 feet could have another tree, and how many quadrants [draws a square with 660 marked on each side and divides the square into quadrants].
So that would be 660 again.
So 660 squared.
So it would be like this, 660 this way and 660 that way, and everyone would have a tree.
Alright. Maybe I can get a job for the forest service [squares 660].

So a square mile would have 435,000 trees and [inaudible].
[Multiplies 435,600 by 1.5 million.]
So there's 6-5-3-9-0-0 million trees.
Oh lord, and how many leaves do they have?
Well that's pretty random.
Let's see, a tree has a lot of branches [draws a tree with branches].
That'd probably have 500 to 1000 leaves on every branch of a tree. 8 to 10.
500, so I'll say every branch has 750 leaves on it.
And there's 8 to 10.
[Multiplies 750 by 9.] 6,750 leaves on a tree.
10,000, I think there's more [increases 6,750 to 10,000 but then crosses it out].

Whoops. [Multiplies 653900 million by 6750.] Alright, I figured it out [laughs].
That's the first part.
I didn't answer, that's a reasonable, that's a reasonable middle estimate. That's a lot of leaves.
So we'd say more would be...

E: I have more paper if you need it.

Oh well I think I'll just refer to my computer.
This is an average.
That's a lot of leaves.
But you know I have, what comes after a million? A billion, then a trillion.
I have 4 trillion million leaves. But that's --

E: How many zero's?

I have 3 zeros on the million. [She has the number written as 4,393,025,000 million.]
So I have, you know like dollar signs.
Like this would be, that would be 9 zeros and then 7 more numbers. That's a lot.
So it's in the trillions of millions.

E: If it makes you feel any better that's very close to the answers everyone else is giving. Which doesn't mean anything.

Well it's this is really good work for the forest service.
Alright, now when I wrote this number 0-0-0 and then million so you have to add on the other million.
You can figure that out.
And then the lower number might be a couple of orders.

What is the question?
Lower estimate. We'll just add another zero.
5-2-8-3-9-3-4 [her answer written backwards] million.
And the lower would be... [Writes 439382500 million.]

E: How much faith do you have in your answer?

None. Okay, I'm done.

Appendix D.

CHOCOLATE PROBLEM:

(1) Now, If either I knew the energy content by some conversion factor, energy content of a chocolate bar or of chocolate generally, I could figure out how much energy there was in that bar.
Then that'd be straight forward, I'd simply say that that energy's converted into potential energy of climbing which would give me MGH for the potential energy.
The mass of person easily estimated, say it's 70 kilograms.
G is 10 in [?] units and H would then be the quantity to be found.
The answer would have to be divided by 5 because of the 20 percent efficiency.
And that would be that.

(2) So the method's fine provided I have a conversion factor, or I have some other comparison.
I seem to have neither.
Is there any other way out?
No.

(3) The only thing I have to estimate, somehow I have to estimate how much enery there is in this chocolate bar.
Any other sources of comparison?
Well maybe I could estimate it if I used one number I do know.
I would have to use approximately 100 joules per second to stay alive.
Okay.

(4) Well I could tell how many joules for a whole day, but how do I compare that to the extra muscular work...
Oh alright.
Okay, let's say that I have, that I guess from my everyday knowledge of food intake that 4 bars would give me life for a day.
So that means 4 bars would last me 24 hours.
So we have 24 hours time 60 minutes times 60 seconds.
And this gives me the number of seconds, so let's call that X seconds.

(5) Oh, an easier way.
Ah, okay, how about this.
4 bars will keep me alive for an entire day with a rate of energy expenditure of 100 joules per second.

(6) Okay, now climbing would be probably 4 times, I probably expend 4 times my energy at an average pace, but 4 times my energy than just staying alive?
Just guessing.
Perhaps, should I make it 5 times more energy than just staying alive?

(7) Okay, so in other words.
Okay here we go.
4 bars keep me alive for a day.
Therefore, that same bar would keep me walking for only a fifth of that, or one and a fifth

hours.
Okay, however there's only 20 percent efficiency, oh wait, no that's fine.
Right.
One bar will keep me climbing for one and a fifth hours.
Alright.

(8) Now as to how high I could climb during that - where does the 20 percent efficiency come in?
Oh, it doesn't come in anymore because I have assumed, I've assumed something else. Simply that I've used 5 times more energy than just staying alive.
Okay which...hum...so in fact I've gone a different way.

(9) So I can climb steadily for one and a fifth hours which is, let's call it 70 minutes, okay. Now as to how high I can climb, I would be walking at, say, 3 kilometers, say, 4 kilometers per hour.
No, walking at 3 kilometers per hour but at the same time going upwards at only about a sixth of that, if I'm lucky.
So that would be up at about half a kilometer every hour.
Okay, and we're going for just about an hour, just over an hour.
So I'd say that I could go up about a half a kilometer, vertically.